MULTICHANNEL THRESHOLDING WITH SENSING DICTIONARIES

Rémi Gribonval, Boris Mailhe, Holger Rauhut, Karin Schnass and Pierre Vandergheynst

Ecole Polytechnique Fédérale de Lausanne (EPFL)
Signal Processing Institute - ITS
CH- 1015 Lausanne, Switzerland

ABSTRACT

This paper shows introduces the use *sensing dictionaries* for p-thresholding, an algorithm to compute simultaneous sparse approximations of multichannel signals over redundant dictionaries. We do both a worst case and average case recovery analyses of this algorithm and show that the latter results in much weaker conditions on the dictionary, sensing dictionary pair. We then do numerical simulations to confirm our theoretical findings, showing that p-thresholding is an interesting low complexity alternative to simultaneous greedy or convex relaxation algorithms for processing sparse multichannel signals with balanced coefficients, and finally point a connection to compressed sensing exploiting the additional freedom in designing the sensing dictionary.

1. OUR PROBLEM AND AN ALGORITHM TO SOLVE IT

Suppose we are to design a network of N sensors monitoring a common phenomenon. Each of our sensors observes a d-dimensional signal $y_n \in \mathbb{R}^d$, $n = 1, ..., N$, where the set of signals obeys a strong sparsity hypothesis, ie. we will assume that each y_n admits a sparse approximation over a single dictionary $\mathbf{\Phi}$:

$$y_n = \mathbf{\Phi} x_n + e_n,\ n = 1, ..., N.$$

In order to model correlations between signals, we will refine this model by imposing that all signals share a common sparse support, i.e.

$$y_n = \mathbf{\Phi}_\Lambda x_n + e_n,$$

where $\mathbf{\Phi}_\Lambda$ is the restriction of the synthesis matrix $\mathbf{\Phi}$ to the columns listed in the set Λ. This model is inspired by a recent series of papers on distributed sensing, see [1] and references therein. It describes a network of sensors monitoring a signal with a strong global component that appears at each node. Localized effects are modeled by letting synthesis coefficients $x_n \in \mathbb{R}^S$, $S := |\Lambda|$, vary across nodes and through the noise e_n. In order to obtain a sufficiently general model, we will assume that the components $x_n(k)$ of the random vector x_n are independent Gaussian variables of variance α_k. This model is fairly general to accommodate various practical problems: the Gaussian assumption is one of the most widely used in signal processing, while incorporating different variances allows us to shape the synthesis coefficients, imposing statistical decay for example on the $x_n(k)$. In order to simplify the analysis we adopt a global matrix notation. We collect all signals y_n on the columns of the $d \times N$ matrix Y and the synthesis coefficients x_n on the columns of the $S \times N$ matrix X. Let U be a $S \times N$ random matrix with independent standard Gaussian entries and let D be a $S \times S$ diagonal matrix whose entries are positive real numbers α_k. Our model can then be written in compact form

$$Y = \mathbf{\Phi}_\Lambda X + E = \mathbf{\Phi}_\Lambda D U + E, \tag{1}$$

where E is a $d \times N$ matrix collecting noise signals e_n on its columns. The problem we will face in this paper is to recover the joint support Λ by sensing the set of signals in a very simple way.

Let us now turn to describing the reconstruction algorithm. The observed signals y_n are sent to a central processing unit that tries to recover the common sparse support Λ. The problem thus boils down to estimating the joint sparse support of a set of signals generated from a redundant dictionary $\mathbf{\Phi}$. A number of algorithms have been proposed lately to jointly process sparse signals, most of them based on multichannel generalizations of greedy algorithms [10] or convex relaxation algorithms. A common weakness to all these techniques is a high computational complexity. To overcome this problem, we would like to resort here to one of the simplest possible algorithms: thresholding. More precisely, our algorithm computes the p-norm of the correlation of the multichannel signal Y with the atoms ψ_k of a sensing dictionary $\mathbf{\Psi}$:

$$\|\psi_k^\star Y\|_p^p := \sum_{n=1}^{N} |\langle \psi_k, y_n \rangle|^p.$$

The sensing dictionary $\mathbf{\Psi}$ has the same cardinality as $\mathbf{\Phi}$, so the atoms in both dictionaries are in a one-to-one relation-

R. Gribonval and B. Mailhe are with IRISA, Rennes, and H. Rauhut with the University of Vienna. This work was completed while they were visiting EPFL and partly supported by the HASSIP network, HPRN-CT-2002-00285.

Contact author: `karin.schnass@epfl.ch`

Index Terms: multidimensional signal processing

ship and corresponding atoms satisfy $\langle \psi_i, \varphi_i \rangle = 1$. Note that because of this normalisation the sensing atoms do not necessarily have unit norm. We could set $\mathbf{\Psi} \equiv \mathbf{\Phi}$, but we as we will see later like this we keep the possibility of optimising both dictionaries in the spirit of [7] or adding additional requirements.

Define Λ_S, the set of indices k with the S largest p-norms. The algorithm is successful if for $S = \sharp\Lambda$ we have $\Lambda_S = \Lambda$. Since $\mathbf{\Psi}^\star Y = \mathbf{\Psi}^\star \mathbf{\Phi}_\Lambda X + \mathbf{\Psi}^\star E$, the strongest p-norm of projections on the set $\overline{\Lambda}$ of bad atoms is

$$\|\mathbf{\Psi}^\star_{\overline{\Lambda}} Y\|_{p,\infty} \leq \|\mathbf{\Psi}^\star_{\overline{\Lambda}} \mathbf{\Phi}_\Lambda X\|_{p,\infty} + \|\mathbf{\Psi}^\star_{\overline{\Lambda}} E\|_{p,\infty},$$

where the (p,∞)-norm of a matrix $\|\mathbf{M}\|_{p,\infty}$ is defined as the maximum of the p-norms of its rows. Conversely, the smallest p-norm of projections on the set of good atoms reads

$$\min_{i\in\Lambda} \|\psi_i^\star Y\|_p \geq \min_{i\in\Lambda} \|\psi_i^\star \mathbf{\Phi}_\Lambda X\|_p - \|\mathbf{\Psi}^\star_\Lambda E\|_{p,\infty}.$$

and the algorithm will thus succeed as soon as

$$\begin{aligned}\min_{i\in\Lambda} \|\psi_i^\star \mathbf{\Phi}_\Lambda X\|_p - \|\mathbf{\Psi}^\star_{\overline{\Lambda}} \mathbf{\Phi}_\Lambda X\|_{p,\infty} > &\|\mathbf{\Psi}^\star_\Lambda E\|_{p,\infty} \\ &+ \|\mathbf{\Psi}^\star_{\overline{\Lambda}} E\|_{p,\infty}.\end{aligned} \tag{2}$$

2. WORST CASE BEHAVIOUR OF P-THRESHOLDING

The recovery condition (2) can be checked based on simple characteristics of the multichannel signals and the dictionaries. To capture the requirements on the dictionary pair we adapt the definition of the standard cumulative coherence [9]:

$$\begin{aligned}\mu_q(\mathbf{\Psi}, \mathbf{\Phi}, \Lambda) &:= \sup_{l\notin\Lambda} \|\mathbf{\Phi}^\star_\Lambda \psi_l\|_q \\ &= \sup_{l\notin\Lambda} \left(\sum_{i\in\Lambda} |\langle \psi_l, \varphi_i \rangle|^q \right)^{1/q}.\end{aligned} \tag{3}$$

As for properties of the signal we need to define the p-Peak SNR and the dynamic range R_p:

$$\begin{aligned}\mathrm{PSNR}_p &:= \frac{\|\mathbf{\Psi}^\star_{\overline{\Lambda}} E\|_{p,\infty} + \|\mathbf{\Psi}^\star_\Lambda E\|_{p,\infty}}{\|X\|_{p,\infty}}, \\ R_p &:= \frac{\min_{i\in\Lambda} \|X(i)\|_p}{\|X\|_{p,\infty}},\end{aligned}$$

where we denote with $\|X(i)\|_p = (\sum_{n=1}^N |x_n(i)|^p)^{1/p}$ the p-norm of the i-th row of X. Following the analysis in [3], it is easy to check that the following condition implies (2):

$$\begin{aligned}\mu_1(\mathbf{\Psi}, \mathbf{\Phi}, \Lambda) + \sup_{i\in\Lambda} \mu_1(\mathbf{\Psi}_\Lambda, \mathbf{\Phi}_\Lambda, \Lambda/\{i\}) \\ < R_p - \mathrm{PSNR}_p.\end{aligned} \tag{4}$$

The success of p-thresholding is thus governed by the condition that the dynamic range of the signal should be bigger than the noise level and the sum of cross correlations among atoms on the support and between the support and the remaining of $\mathbf{\Phi}$. We note that μ_1 can be very big even for reasonably small Λ. For example, when $\mathbf{\Psi} = \mathbf{\Phi}$, the quantity $\mu_1(\mathbf{\Psi}, \mathbf{\Phi}, \Lambda) + \mu_1(\mathbf{\Psi}_\Lambda, \mathbf{\Phi}_\Lambda, \Lambda/\{i\})$ is often replaced by its upper estimate $(2S-1)\mu$. The r.h.s in (4) is at most one, so the resulting condition can only be satisfied when $S < (1+\mu^{-1})/2$. In the next sections, we develop an average case analysis of p-thresholding and show that the *typical* recovery conditions are much less restrictive.

The first contribution of this paper is to show when the simple p-thresholding algorithm using the sensing dictionary $\mathbf{\Psi}$ will succeed in recovering the correct support Λ with high probability. As we will see below, the sparsity constrain is expressed in terms of the 2-cumulative coherence μ_2 and is thus much weaker than worst case conditions that are usually expressed in terms of μ_1. Moreover, the recovery probability scales exponentially with the number of channels.

3. AVERAGE CASE ANALYSIS OF P-THRESHOLDING

To state the central theoretical result for the average case we need to define a probabilistic PSNR and dynamic range, remember we had $Y = \mathbf{\Phi}_\Lambda D U + E$ where $D = \mathbf{diag}(\alpha_i)$,

$$\begin{aligned}\overline{\mathrm{PSNR}}_p &:= \frac{\|\mathbf{\Psi}^\star_{\overline{\Lambda}} E\|_{p,\infty} + \|\mathbf{\Psi}^\star_\Lambda E\|_{p,\infty}}{\max_{i\in\Lambda} |\alpha_i|}, \\ \overline{R} &:= \frac{\min_{i\in\Lambda} |\alpha_i|}{\max_{i\in\Lambda} |\alpha_i|}.\end{aligned}$$

Theorem 1. *Assume that the noise level and the dynamic range are sufficiently small (respectively large), that is to say*

$$\mu_2(\mathbf{\Phi}, \mathbf{\Psi}, \Lambda) < \min_{i\in\Lambda} \|\mathbf{\Phi}^\star_\Lambda \psi_i\|_2 \cdot \overline{R} - \overline{\mathrm{PSNR}}_p / C_p(N). \tag{5}$$

where $C_p(N)$ is a constant depending only on p and the number of channels N. Then, under signal model (1), the probability that p-thresholding fails to recover the indices of the atoms in Λ does not exceed

$$\mathbb{P}(p - \textit{thresholding fails}) \leq K \cdot \exp\left(-A_p(N)\gamma^2\right)$$

with

$$\gamma = \frac{\overline{R} \cdot \min_{i\in\Lambda} \|\mathbf{\Phi}^\star_\Lambda \psi_i\|_2 - \overline{\mathrm{PSNR}}_p / C_p(N) - \mu_2(\mathbf{\Phi}, \mathbf{\Psi}, \Lambda)}{\overline{R} \cdot \min_{i\in\Lambda} \|\mathbf{\Phi}^\star_\Lambda \psi_i\|_2 + \mu_2(\mathbf{\Phi}, \mathbf{\Psi}, \Lambda)}.$$

and $A_p(N)$ is a constant depending on p and growing with the number of channels N, e.g $A_1(N) = N/\pi$, $A_2(N) \sim N/2$.

The proof of this result is somewhat lengthy and relies heavily on measure concentration inequalities. The interested reader will find all details in [4]. This result has unique features compared to more classical worst case analysis. First, the condition on the dictionary pair is expressed in terms of the cumulative coherence of order 2 which is much smaller than that of order one. For example assuming that there is no noise and that the variances α_i are constant the r.h.s in (5) is larger than one. If additionally $\mathbf{\Psi} = \mathbf{\Phi}$, an upper estimate of $\mu_2(\mathbf{\Phi}, \mathbf{\Psi}, \Lambda)$ is $\mu\sqrt{S}$ and we see that typically thresholding can be successful even when $S \approx \mu^{-2} \gg \mu^{-1}$. Second, due to typicality, we see that the probability of failure quickly diminishes as the number of channel grows, suggesting that we should use $N \sim \log K$ channels in practice. These findings are confirmed by simulation results as we show in the next section.

4. EXPERIMENTAL RESULTS

In this section we compare our theoretical findings with simulations of the performance of 2-thresholding with $\mathbf{\Psi} = \mathbf{\Phi}$. As dictionary we chose a combination of the Dirac and Fourier basis, $\mathbf{\Phi} = (\mathbf{I}_d, \mathcal{F}_d)$, in dimension $d = 1024$, which has coherence $\mu = 1/\sqrt{d}$. For each number of channels N, varying from 1 to 128, and support size, varying from 1 to 1024 in steps of 16, we created 180 signals by choosing a support Λ uniformly at random and independent Gaussian coefficients with variances $\alpha_i = 1$ and calculated the percentage of thresholding being able to recover the full support. The results can be seen in Figure 1.

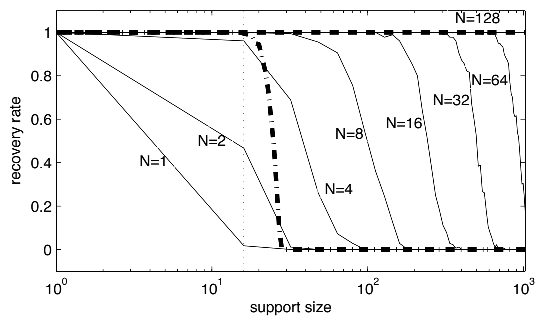

Fig. 1. Comparison of Recovery Rates for Different Support Sizes and Number of Channels.

As reference we also calculated how many out of 200 randomly chosen supports of a given size satisfy the worst case recovery condition $\mu_1(\lambda) + \sup_{i\in\Lambda} \mu_1(\Lambda/\{i\}) < 1$. This is indicated by the dash dotted line and can be seen to drop rapidly once the theoretical limit $|\Lambda| = 16$ is reached. Since $\mu = 1/\sqrt{d}$ the average recovery condition $\mu_2(\Lambda) < 1$, indicated by the dashed line, is always satisfied. We can see that as predicted by Theorem 1 with an increasing number of channels we get closer to the average case bound, which is actually attained once $N = 128$.

5. SENSING DICTIONARY DESIGN

As we have seen in the last section the success of p-thresholding is largely determined by the order 2 cross-correlation between the original and the sensing dictionary. Thus in order to optimise the performance of the algorithm we should choose the sensing dictionary the minimises the order 2 cross correlation, ie.

$$\mathbf{\Psi}_{opt} = \arg\min_{\mathbf{\Psi}} \mu_2(\mathbf{\Phi}, \mathbf{\Psi}).$$

The same optimisation problem arises when studying the average performance of the simple thresholding algorithm in [6], which we refer to for more details.

Instead we will point out an interesting connection between compressed sensing and sensing dictionaries of low rank. We introduced our signal model with the example of having to design a network of N sensors monitoring a common phenomenon, where each of the sensors sends his observation to a common processing unit, which then finds the common support, which we were interested in. The disadvantage in this scheme is that every sensor needs to obtain and send the whole signal, which is both time and energy consuming. So if we are already close to performance break-down for the given number channels and support size there is nothing we can do. However assume that we have a lot of channels or very small support size and that the dictionary is very well behaved, ie. it is easy to find a very good sensing dictionary. In this case we could try to find a sensing dictionary which still gives reasonable cross-coherence but of rank much lower than the dimensionality of the signals :

$$\mathbf{\Psi}_{opt} = \arg\min_{\mu_2(\mathbf{\Phi},\mathbf{\Psi})<c} \operatorname{rank}(\mathbf{\Psi}).$$

The advantage of such a low rank sensing dictionary is that we can reduce the size of the signal that every sensor has to pass on. So assume that $\mathbf{\Psi}$ has rank q and thus has a reduced singular value decomposition of the form

$$\mathbf{\Psi} = \underset{d\times d}{U} \cdot \underset{d\times q}{\Sigma} \cdot \underset{p\times K}{V^\star} \tag{6}$$

where both U, V have orthonormal columns. Multiplying the signals by $\mathbf{\Psi}^\star$ in the algorithm amounts to first multiplying by $\Sigma^\star U^\star$ and then by V. However if we do the first multiplication not at the fusion center but at the sensors only a signal of size q instead of d has to be send on. In order to get a feeling for how much reduction is possible, let's adopt an alternative viewpoint. By multiplying with $\Sigma^\star U^\star$ we get signals Z that are sparse in the dictionary $\mathbf{\Phi}^q := \Sigma^\star U^\star \mathbf{\Phi}$, ie

$$Z = \Sigma^\star U^\star \Phi_\Lambda X + \Sigma^\star U^\star E = \mathbf{\Phi}^p_\Lambda + \tilde{E}$$

which we want to reconstruct using the sensing dictionary $\mathbf{\Psi}^q := V^\star$. Since the dictionaries have K atoms in dimension p the cumulative coherence will be of the magnitude $\mu_2(\mathbf{\Phi}^q, \mathbf{\Psi}^q, S) = \sqrt{S/q}$. If we insert this estimate into the formula for γ from Theorem 1, assuming $\bar{R} = 1, E = 0$ for simplicity. We can estimate that this kind of compressed 1-thresholding fails by

$$\mathbb{P}(1\text{−thresholding fails}) \leq K \cdot \exp\left(-\frac{N}{\pi}\left(\frac{1-\sqrt{S/p}}{1+\sqrt{S/p}}\right)^2.\right)$$

Further simplifying the above formula we get as rule of thumb on how compressed we can sense our signals depending on their expected sparsity, dictionary size and number of channels

$$q \geq S\left(\frac{1}{2} - \frac{\log(\epsilon/K)}{AN}\right)^2$$

For a more detailed exposition and an algorithm to calculate low rank sensing matrices we refer to [8].

6. CONCLUSIONS

Thresholding is a computationally inexpensive algorithm for simultaneous sparse signal approximation. We have shown that, in a probabilistic multichannel setting, it shares good recovery properties with much more complex alternatives such as greedy algorithms and convex relaxation algorithms. The worst case recovery condition is reminiscent of Tropp's recovery condition, see [9], but the typical behaviour is instead driven by a much less restrictive condition and improves with the numbers of channels. This is clearly confirmed by our simulation results.

It has to be noted that the results obtained in this paper do not *scale down* to a single channel. Indeed, our average case results rely heavily on typicality across channels. On the other hand, single channel average case results have been obtained for the simple thresholding algorithm in [6] and confirm that the 2-coherence is a characteristic performance measure.

One of the main drawbacks of thresholding is that its performance relies heavily on the assumption that the signal coefficients are well balanced, in addition to the Gaussian model. Orthogonal Matching Pursuit is a natural candidate for dealing with signals that do not have balanced coefficients. Preliminary results [4] indicate that its typical performance in a multi-channel probabilistic setup is also driven by much less restrictive conditions on the dictionary than the worst case ones.

In this paper we have studied a generalized version of the algorithm, allowing ourselves to use a special sensing dictionary for computing projections. We have shown that the average performance of thresholding involves the mutual coherence of order 2 between the sensing and synthesis dictionaries. This could lead to designing sensing dictionaries to optimize the recovery performance for a given signal model. Another interesting question would be to study how practical thresholding can be in the framework of Compressed Sensing [2]. It has been proved in [5] that thresholding can be used a recovery algorithm in this setting and its lower computational complexity (as compared with OMP) might be useful in particular applications. As a first step in this direction, we investigated a stylized distributed Compressed Sensing scenario and shown that the freedom of designing an optimized sensing dictionary could potentially strongly reduce the number of projections computed at each node.

7. REFERENCES

[1] D. Baron, M.F. Duarte, S. Sarvotham, M.B. Wakin, and R.G. Baraniuk. An information-theoretic approach to dis- tributed compressed sensing. In *Proc. 45rd Conference on Communication, Control, and Computing*, 2005.

[2] E. Candes, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59:1207–1223, 2005.

[3] R. Gribonval, M. Nielsen, and P. Vandergheynst. Towards an adaptive computational strategy for sparse signal approximation. IRISA preprint, Rennes, France, 2006.

[4] R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst. Atoms of all channels, unite! Average case analysis of multi-channel sparse recovery using greedy algorithms. Technical Report PI-1848, IRISA, 2007.

[5] H. Rauhut, K. Schnass, and P. Vandergheynst. Compressed sensing and redundant dictionaries. *submitted to IEEE Trans. Inform. Theory*, 2006.

[6] K. Schnass and P. Vandergheynst. Average performance analysis for thresholding. *accepted to IEEE Signal Processing Letters*, 2007.

[7] K. Schnass and P. Vandergheynst. Dictionary preconditioning for greedy algorithms. *accepted to IEEE Trans. Signal Processing*, 2007.

[8] K. Schnass and P. Vandergheynst. Dimensionality reduction for classification. Technical report, EPFL-ITS, 2007.

[9] J. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Transactions on Information Theory*, 50(10):2231–2242, October 2004.

[10] J.A. Tropp, A.C. Gilbert, and M.J. Strauss. Algorithms for simultaneous sparse approximations. part i: Greedy pursuit. *Signal Processing, special issue "Sparse approximations in signal and image processing"*, 86(572-588), 2006.